

Reto - Interpolación y ajuste de curvas

David Ramirez
Ingeniería de Sistemas
Pontificia universidad javeriana.
Bogotá, Colombia
da-ramirezm@javeriana.edu.co

Andrés Giraldo
Ingeniería de Sistemas
Pontificia universidad javeriana.
Bogotá, Colombia
giraldoma@javeriana.edu.co

Camilo García
Ingeniería de Sistemas
Pontificia universidad javeriana.
Bogotá, Colombia
candres-garcia@javeriana.edu.co

Resumen—En el presente documento se pretende describir la metodología, métodos numéricos y ajustes lineales utilizados en el desarrollo del reto número dos, con el fin de dar a entender su importancia y funcionamiento dentro de la implementación desarrollada.

Index Terms—interpolación, python, ajuste de curvas

I. INTRODUCCIÓN

Dentro del desarrollo propuesto para la elaboración del reto, se hace explícita la necesidad de métodos interpolantes para la obtención de nuevos valores según un grupo de datos establecido. En general el entorno en donde se pretende aplicar dicha teoría gira con base en los datos de diferentes estaciones meteorológicas a lo largo de la región brasileña, estos funcionan como variables ajustables que permiten la obtención y aproximación de sus valores en otras estaciones meteorológicas. En términos generales se pretende generar una interpolación dado este grupo de datos, y analizar las coincidencias e índices de similitud definidos más adelante con el fin de comprobar la teoría matemática detrás de estos métodos numéricos.

II. DESCRIPCIÓN

El contexto general del reto, como se dijo anteriormente, se presenta en forma de datos obtenidos por diferentes estaciones meteorológicas a lo largo de brasil, con los cuales se pretende realizar una comparativa aleatoria sobre el 70 por ciento de los datos interpolados tomando como referencia una variable (temperatura) dentro de cierta estación, generando gráficas comparativas sobre el comportamientos de esta variable a través del tiempo y permitiendo realizar una comprobación de efectividad del método de interpolación con el 30 por ciento de los valores restantes sobre los valores ya almacenados por la estación meteorológica escogida.

Como segundo caso, también, se pretende utilizar la ubicación geográfica de cada una de las estaciones con el fin de encontrar su estación más cercana, para así a través de varios métodos interpolantes predecir los datos de dicha estación con base en los datos de su estación más cercana, y generar un índice de similitud entre estos 2 conjuntos de datos, analizando la efectividad de la interpolación y dando paso al análisis de dichos datos de forma más profunda.

Identify applicable funding agency here. If none, delete this.

III. MARCO TEÓRICO

Para la solución de los ejercicios propuestos y el desarrollo del reto general se requirió la implementación de varios algoritmos matemáticos que se mostraran a continuación.

III-A. Splines cúbicos

Los splines cúbicos se definen como una gran cantidad de funciones que son utilizadas en aplicaciones que requieren interpolación de datos y ajuste de curvas, dadas por la función de grado 3.

$$p(x) = ax^3 + bx^2 + cx + d = 0$$

III-B. Interpolación Cuadrática.

La interpolación cuadrática en general consiste en hallar un dato entre los intervalos que se conocen como extremos, al ser interpolación cuadrática se deben tomar 3 puntos para determinar el intervalo, así mismo si la interpolación fuese lineal se tomaron solo 2 puntos.

$$p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

III-C. Indice de jaccard

El índice de jaccard se define como una medida de similitud entre dos conjuntos, o más formalmente como el tamaño de la intersección sobre el tamaño de la unión.

$$J(A, B) = \left| \frac{A \cap B}{A \cup B} \right|$$

III-D. Error cuadrático medio

El error cuadrático medio se puede definir como el error que existe dentro de dos conjuntos de datos, comparando el valor aproximado con el valor real, también puede denotarse como la raíz de la desviación cuadrática media.

$$EMC = \frac{\sqrt{\sum_{i=1}^n (Z^*(x) - Z(x))^2}}{n}$$

IV. PRESENTACIÓN DE LOS DATOS

Para el desarrollo del reto fue proporcionada una base de datos en un documento excel. En dicho documento, se presentan las variables climáticas de diecisiete estaciones, distribuidas a lo largo del territorio de Brasil. Entre estas variables se pueden encontrar:

- Año de los datos.
- Hora (militar).

- Día Juliano.
- Temp Interna (C°).
- Presión atmosférica.

Adicionalmente, se realizó una tabla con la ayuda de una imagen con las coordenadas cartesianas correspondientes a cada estación, y que será presentada a continuación.

Estación	x	y
Viçosa	2.7	9.61
Sobral	3.5	9.59
São Gonçalo do Amarante	5	9.6
Santa Quitéria	3.57	9.52
Santana do Cariri	4.2	9.2
Quixeramobim	4.7	9.42
Quixadá	4.7	9.46
Pentecoste	4.7	9.58
Jatí	5	9.16
Itatira	4.3	9.5
Funceme	5.5	9.58
Crato	4.6	9.2
Camocim	3	9.68
Acarau	3.8	9.68
Araípe	3.7	9.2
Beberibe	6	9.52

Figura 1. Tabla coordenadas

V. METODOLOGÍA

Dado que se requiere un método de verificación y comparación de datos para cada uno de los metodos de interpolacion dados, se decidió comparar el comportamiento de cada una de las variables interpoladas respecto a su tiempo correspondiente, tambien asi comparando dicho comportamiento con los valores reales dados en la base de datos consultada.

Primeramente, se obtuvo una muestra aleatoria del primer grupo de datos igual al 70 por ciento, posteriormente se realizó tanto la interpolación cuadrática como los splines cúbicos con el fin de comparar estos dos métodos, obteniendo como resultado valores extremadamente similares. Conclusión que es posible corroborar gracias a el EMC dado, con un valor de más o menos 1 , el cual es considerablemente aceptable para el rango de valores que puede tomar la variable temperatura.

Una vez realizado el primer punto se requiere obtener la aproximación de la variable temperatura de la estación más cercana a otra situación dada, para esto igualmente se realizó la interpolación de la variable temperatura y se realizaron las comparaciones correspondientes con los datos obtenidos. A continuación, se describe el procedimiento realizado.

V-A. Punto 1

- Se selecciona una variable aleatoriamente
- Se selecciona una estación meteorológica
- Se realiza un muestreo del 70 por ciento de los datos
- Se aplican los splines cúbicos con el fin de graficar los datos
- Se aplica la interpolación cuadrática con el fin de graficar los datos

- Se toma el 30 por ciento restante como valores de validación
- Se interpola el 30 por ciento y se compara con los valores reales
- Se gráfica con el fin de comparar cada uno de los métodos de interpolación

V-B. Punto 2

- Se selecciona una variable aleatoriamente
- Se selecciona una estación meteorológica
- Se encuentra la estación más cercana a dicha estación
- Se interpolan los datos de la estación más cercana
- Se comparan los datos reales de la estación con los datos interpolados de la estación mas cercana
- Se grafica tanto la interpolación como los datos reales con el fin de comparar

VI. TABLAS Y GRÁFICOS

VI-A. Primer punto

Para este primer punto se obtuvieron cuatro tres gráficas, para demostrar como se ven afectados los datos de la temperatura al momento de hacerles alguna interpolación (Para este caso Splines cúbicos e interpolación cuadrática). Para identificar si hubo algún cambio significativo se comparan las graficas de los datos interpolados, con los datos originales brindados por el archivo de excel.

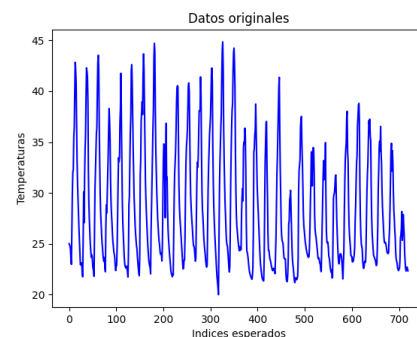


Figura 2. Datos originales

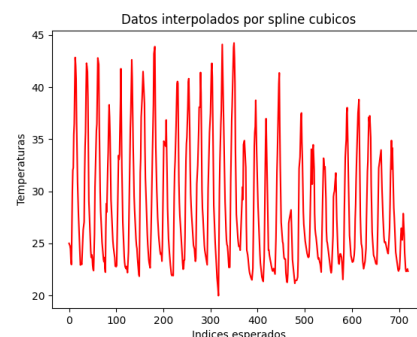


Figura 3. Datos interpolados por medio de splines cubicos

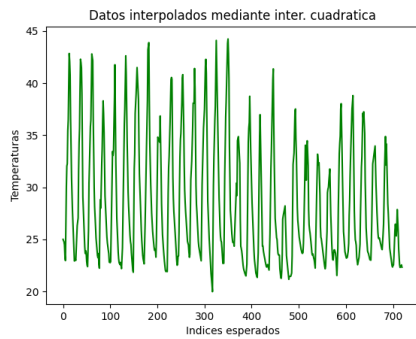


Figura 4. Datos interpolados mediante interpolación cuadrática

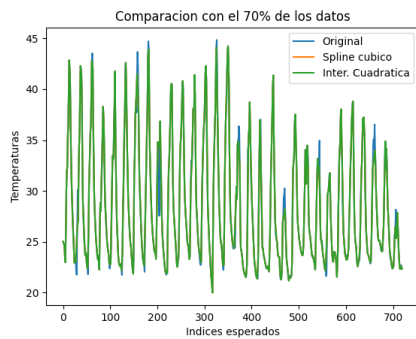


Figura 5. Comparación del 70 % de los datos

Punto 1		
Tipo de Error	Spline Cúbico	Interpolación Cuadratica
EMC	0.08°	0.08°
Indice de Jaccard	0.46	0.46

Figura 6. Tabla de errores punto 1

VI-B. Segundo punto

Para el desarrollo de esta parte del reto, se comparó el resultado de la interpolación de dos estaciones cercanas a la estación de Itatira. Con el fin de determinar si era posible estimar los datos de Itatira a partir de las estaciones Santa Quitéria y Sobral.

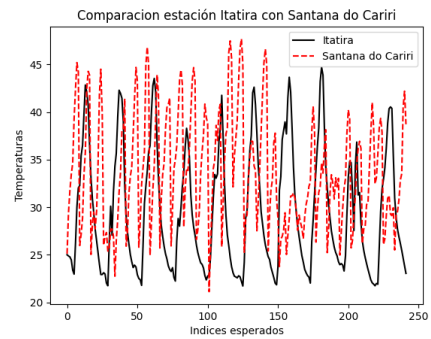


Figura 7. Comparación Itatira con Santana do Cariri

Itatira a Santana do Cariri		
Tipo de Error	Spline Cúbico	Interpolación Cuadratica
EMC	0.65°	0.65°
Indice de Jaccard	0.38	0.38

Figura 8. Errores Itatira - Santana do Cariri

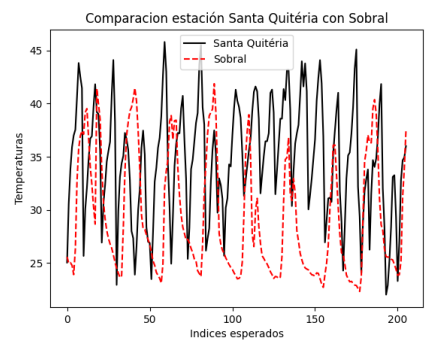


Figura 9. Comparación estación Santa Quitéria con Sobral

Santa Quitéria a Sobral		
Tipo de Error	Spline Cúbico	Interpolación Cuadratica
EMC	0.68°	0.68°
Indice de Jaccard	0.32	0.32

Figura 10. Errores Santa Quitéria - Sobral

VII. CONCLUSIONES

- A pesar que los errores tienen resultados similares, estos varían en un cantidad reducida de cifras decimales.
- La interpolación cuadrática tiende a ser más precisa que los splines cúbicos, ya que esta primera utiliza una mayor cantidad de datos, pero esto le quita eficiencia computacional al método.
-

REFERENCIAS

- [1] Ojeda, L. (2016). Análisis numérico básico. Escuela Superior Politécnica del Litoral.
- [2] Similitud de JACCARD en python. estatologos. OBTENIDO DE: Cómo calcular la similitud de Jaccard en Python en 2021 → VACANCEENEUROPE® (jaol.net)
- [3] P.M. Prenter, "Splines and variational methods", Wiley (1975)
- [4] K. Elissa, "Title of paper if known," unpublished.